

JRC TECHNICAL REPORTS

Report on a demonstrator for the fight against counterfeiting

Classification and verification of consumer electronic devices for fight against counterfeiting: the case study of the mobile phone

Gianmarco Baldini, Raimondo Giuliani



This publication is a Technical report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication.

Contact Information

Name: Gianmarco Baldini

Address: Joint Research Centre, Via Enrico Fermi 2749, TP 360, 21027 Ispra (VA), Italy

E-mail: gianmarco.baldini@jrc.ec.europa.eu

Tel.: +39 0332 78 6618

JRC Science Hub

<https://ec.europa.eu/jrc>

JRC103738

EUR 28388 EN

PDF	ISBN 978-92-79-64829-8	ISSN 1831-9424	doi:10.2788/86841
Print	ISBN 978-92-79-64830-4	ISSN 1018-5593	doi:10.2788/51796

Luxembourg: Publications Office of the European Union, 2016

© European Union, 2016

The reuse of the document is authorised, provided the source is acknowledged and the original meaning or message of the texts are not distorted. The European Commission shall not be held liable for any consequences stemming from the reuse.

How to cite: Gianmarco Baldini et al. Report on a demonstrator for the fight against counterfeiting. EUR 28388 EN. doi 10.2788/86841

All images © European Union 2016

Report on a demonstrator for the fight against counterfeiting

This report provides an overview of the activities carried on in 2016 in the JRC.E in the context of GNSS for road transportation.

Report on a demonstrator for the fight against counterfeiting

Gianmarco Baldini, Raimondo Giuliani

Contents

Executive Summary	1
1. Introduction	2
2. The concept of radio frequency fingerprinting	4
3. Description of the Global System for Mobile Communications (GSM) standard	6
4. Methodology, statistical features and machine learning	7
4.1. Overall workflow	7
4.2. Experimental setup	9
4.3. Statistical features	11
4.4. Machine Learning algorithm.....	12
5. Experimental results	14
6. Conclusions	15
References	16
List of abbreviations and definitions	17
List of figures.....	18

Executive Summary

This report has the objective to describe the demonstrator developed in the context of the project *Goods Digital Identification against Fraud and Counterfeiting*, which supports the European Union Intellectual Property Office (EUIPO). One of the goals of this project is to investigate and develop techniques to detect counterfeit products and electronic devices in particular. Because counterfeit electronic devices are usually built with different (e.g., cheaper) materials or different manufacturing processes than proper electronic devices, one of the main techniques is based on the capability to identify electronic devices on the basis of their physical properties. These physical properties can appear in the digital information generated by the mobile phone in its operation. For example, the Radio Frequency (RF) emissions from the communication systems of a mobile phone can be used to identify the phone and its model. In this report, we describe a demonstrator to identify mobile phones on the basis of the radio frequency emissions generated by the transmission components implemented for the GSM wireless standard. We demonstrate that we can distinguish mobile phones not only for different brands and models but also for phones of the same brand and model and different serial numbers. The demonstrator is designed and implemented using low cost systems and machine learning algorithms. In addition to the fight against counterfeiting, the identification of electronic devices and mobile phones in particular can be used to support forensics and criminal investigation. In this context, this work can also be used to support EUROPOL.

1. Introduction

As described in (1) and (2), the authentication of an electronic device, component or system is an important function in the fight against counterfeiting and Intellectual Property Rights (IPR) infringement in the electronics market. The authors in (1) have defined a taxonomy of the counterfeit Integrated Circuits (IC)s in different categories where the device identification methods described in this report can be applied. These categories include recycled, remarked, out-of-spec/defective or overproduced components. Recycled components are used in IC components, which are repackaged and remarked, and then sold in the market as new. Because they are old and used, recycled components could have a different fingerprints than brand new components due to time wear or degradation. The remarking process includes the removal of markings on the package (or even on the die) and remarking with forged information. The reason for remarking is to obtain an higher specification (e.g., from commercial grade part to industrial or defense grade) and resell a cheaper component for an higher price. In this case, the difference in quality between high specification and low specification components could be identified through the techniques described in this article. For examples, the authors in (3) shows that clock stability in oscillators is directly related to the quality of hardware components. In another example, the authors of (4) have shown that a low grade RF amplifier has a distinct RF signature in comparison to an high grade amplifier and this can be detected by the analysis of the spectral response. Still further research work and studies are needed to evaluate how recycled and remarked electronic components produce different fingerprints from newly produced components. Out-of-spec or defective products can be subject to a similar analysis of remarked products because they are built with components out of specifications or even defective. A key element for the identification of counterfeit electronic components in this category is the knowledge on how a defect can modify the fingerprint of the component. Finally, the overproduced components could be the more difficult to identify as overproducing often means that components are produced in the same foundry and similar materials of proper components outside the contract. In this case, the counterfeiters gain is due to the infringement of the Intellectual Property Rights rather than use cheaper components and materials. As a consequence, the fingerprints due to material or features of the manufacturing line would not be usable to detect counterfeit products. To summarize, the fight against the distribution of counterfeit products can exploit the techniques described in this report for identification purposes. The challenging goal is to find features and algorithms which can distinguish mobile phones on the basis of the physical properties of their RF components. Both supervised and unsupervised machine learning algorithms could be used for this purpose. The first set of algorithms could be used to identify if a mobile phone is a non-counterfeit item on the basis of a previously created reference library of proper phones, while the second set of algorithms could be used to generate clusters of proper and counterfeit phones to support the identification (e.g., on the basis of the similarity to one of the clusters) of new mobile phones to be identified.

In this report, we focus on the application of supervised algorithms, where a reference library based on the RF emissions of known phones is used to identify phones. Counterfeit phones are not available because of strong restrictions by the customs officers in the various European countries. Customs officers are bound to sequester and lock away counterfeit products including mobile phones. Then, in this report a counterfeit phone is simulated by claiming that a phone is of a different model than what it is and by checking if the algorithm is able to verify the claim. This is a common approach used in research literature as demonstrated in (5) and (6). The work presented in this report is part of a wider project for the identification of electronic devices and mobile phones in support of EUIPO and its Observatory. The Observatory is a platform-based body that brings public and private sector experts together in a dynamic exchange network. The Observatory works as a think-tank that gathers, monitors and reports crucial information to assist policy makers and authorities engaged in protecting and enforcing IP rights (see <https://euipo.europa.eu/ohimportal/en/european-observatory>). In addition to the fight against counterfeiting, the identification of electronic devices and mobile phones in particular can be used to support forensics and criminal investigation. In this context, this work can also be used to support EUROPOL.

The structure of the report is following: section 2 describes the main concepts of RF fingerprinting. Because GSM is the standard of choice for fingerprinting used in this report, a description of the GSM standard is provided in section 3. Then, section 4 describes the methodology and the application of statistical features and the machine learning algorithms used to classify and verify the mobile phones. The results of the application of the algorithms are described in section 5. Finally, section 6 provides the conclusion of the report.

2. The concept of radio frequency fingerprinting

In this context, we use the term *fingerprinting* to refer to the process by which observable characteristics are extracted from a mobile phone in order to make it identifiable and distinguishable from another one of the same brand or even of the same model. The observation of these characteristics can be performed in different ways, which will be described in this report for the specific case of the GSM radio frequency built-in components of a mobile phone. These fingerprints are usually generated in the preparation of the base materials of the components and in the manufacturing process, and their insertion is accidental or intrinsic to the process itself. However, they can also be inserted on purpose like, for example, the Physical Unclonable Functions (PUF) concept where physical entities are embodied on purpose in the physical structure of a component (7). An example of a PUF is the insertion of specific electronic circuits in an electronic device to generate a reproducible delay in a specific operation executed by the electronic device. In both cases, fingerprints are usually tiny variations in the electronic components which can be exploited for the identification of a mobile phone model or the phone itself if they can generate observable characteristics, which can be collected and analyzed with an adequate level of precision. The term adequate is relative to the type of imperfections, the way the observables are collected and other factors. For example, if the type of imperfections does not appear clearly in the observables or if their value is depending on environment factors (e.g., temperature), these imperfections cannot be used to generate the fingerprints because they will not be reproducible in all the conditions. In another example, the observables must be collected and processed by a system with a high level of precision to distinguish the small differences in the observables. This can be seen in the processing of RF signals performed by two different test receivers. The level of accuracy of a high precision receiver can be much higher than a low quality receiver (see (8)). There is obviously a trade-off because the improvement in accuracy is gained with an increased cost of the receivers. The fingerprinting of electronic components has many similarities to the fingerprinting of human beings in biometrics. Indeed, some requirements for fingerprinting defined in the biometrics domain (9) and (10) can also be adopted for the fingerprinting of mobile phones:

1. universality, which means that every mobile phone or its electronic components should have the characteristics that are used for identification;
2. uniqueness, which indicates that no two components should have the same fingerprinting or physical characteristics;
3. permanence, which means that the characteristic should be invariant with time or with the environment conditions;
4. collectability, which indicates that the characteristics can be measured quantitatively.

In the literature, the terms identification, classification, authentication and verification are sometimes used with slightly different meanings. For this reason, we provide the following definitions, which we will be used in the rest of this report:

1. Authentication: is the process of actually confirming the claimed identity of a phone. Most of the techniques described in this paper have the objective to authenticate a phone through the physical fingerprints of their components, which are difficult or impossible to clone.
2. Verification: is a synonym of authentication in this context as it verifies the claimed identity of a phone.
3. Identification: the recognition system determines a device identity by comparing a captured device fingerprint with reference fingerprint templates for all known devices. Identification requires a one-to-many comparison and is considered more difficult than verification. Note that identification in this context has a different meaning than in other contexts where identification is the process by which an entity profess to have a certain identity.
4. Classification: is the process by which mobile phones are classified in different classes or categories.

Verification and authentication is usually performed with a binary classifier: the observables of two mobile phones are compared using a machine learning algorithm, while identification and classification is usually performed with a multi-classifier algorithm. The definition of the algorithms will be described in the section 4.

3. Description of the GSM standard

GSM is a well established standard in digital communications developed by the European Telecommunications Standards Institute (ETSI), describing protocols for second-generation (2G) digital cellular networks.

GSM is the widest deployed wireless communication standard in the world, with a pre-dominant customer base, even if 3G and 4G digital cellular networks have already been, or are just to be, deployed in a variety of countries. There are billions of GSM mobile phones in the world, which makes their distribution quite widespread and the cost minimal. Then, we can assume that most of the users have a GSM mobile phone.

In this section, we focus on the description of the air interface of the GSM standard, because for the identification of the GSM phones, the understanding of the physical layer of the GSM standards is the most important.

GSM utilizes a combination of Frequency Division Multiple Access (FDMA) and Time Division Multiple Access (TDMA) on the Air-interface (11). FDMA and TDMA are used in combination to produce a two-dimensional array of atomic elements, which are called the Time slots (TS). In full-rate configuration, eight TS are mapped on every frequency.

In a GSM system, every TDMA frame is assigned a fixed number, which repeats itself in a time period of 3 hours, 28 minutes, 53 seconds, and 760 milliseconds. This time period is referred to as hyperframe. The hyperframe is then structured in multiframes, whose number is dependent on the specific configuration in traffic channels or signaling channels. The details on the GSM standard can be found in (11).

The structure of the GSM standard at the physical layer is shown in figure 1.

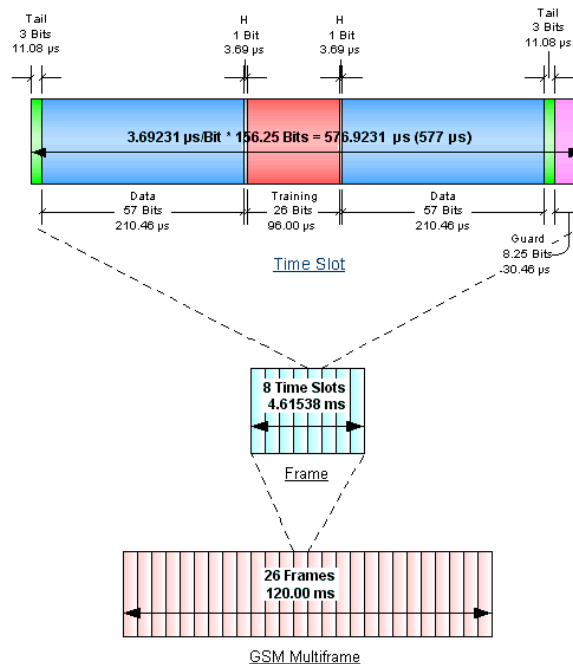


Figure 1: GSM burst definition from (11)

Each time slot lasts for approximately 577 microseconds. The TS are defined in the same mode both for the GSM uplink channel (i.e., from the GSM mobile phone to the Base Station) and the GSM downlink channel (i.e., from the Base Station to the GSM mobile phone). In this report, we use the TS of the GSM uplink channel because we want to locate the GSM mobile phone rather than the GSM base station (which is known).

Each burst is transmitted independently by the GSM mobile phone, so the procedure of sending a *normal burst* is always the same for the RF hardware of the GSM mobile, while the content (e.g., phone conversation) carried in the payload in the traffic channels may be different.

A practical (e.g., sample of a GSM communication) representation in the time domain of a GSM burst, as we will call it in the rest of this report is presented in figure 2. The structure of the burst from figure 2 is following:

1. At the start of a burst, the sending power ramps up to the desired signal strength. This ramp-up section is independent from the content. This is identified by shape 1 in the figure of the GSM burst.
2. The data payload is split into two data segments of 57 Bits each, with a training sequence in between. The data segments are content-related and their use for the fingerprinting of the GSM mobile phone is difficult because the content would change from one burst to another. The content related sections are identified by shapes 2a and 2b in the figure of the GSM burst.
3. The midamble or training sequence (shape 3) is not content related and it can be used for the fingerprinting of the GSM mobile phone.

To summarize, the fingerprinting can be based on all the non-content shapes of the GSM burst, which include the ramp up, the ramp down and the training sequence. The ramp up and the ramp down could be sensible to attenuation and multi-path fading effects and they are usually characterized by a limited number of samples.

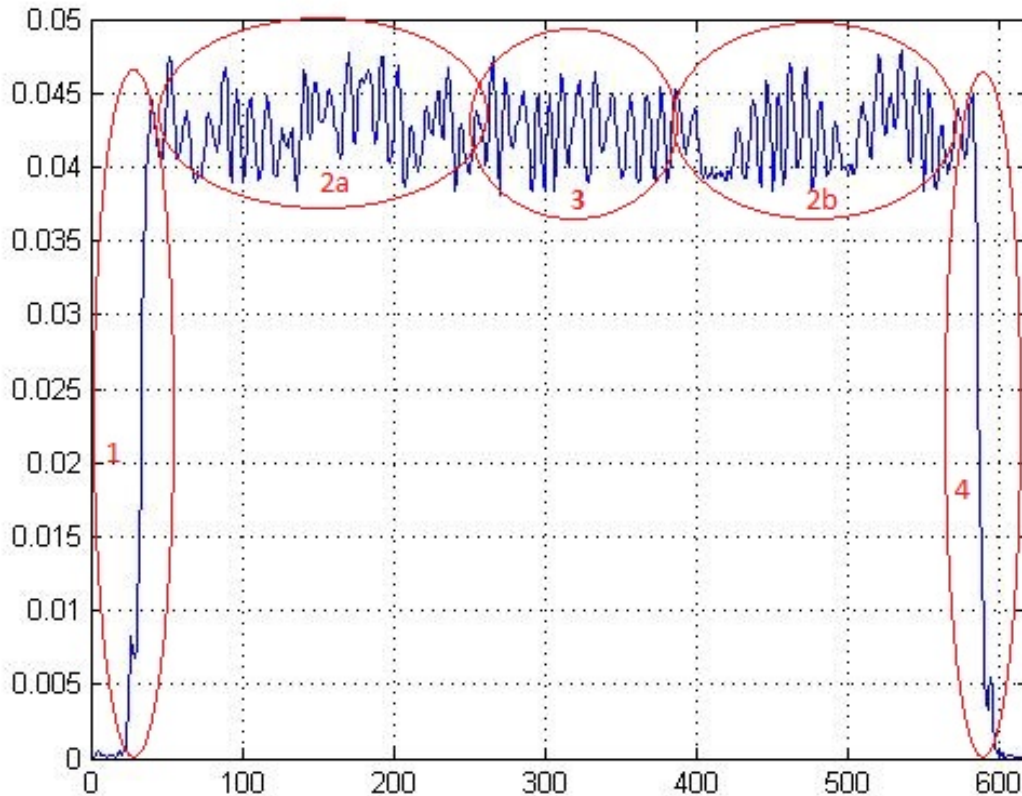


Figure 2: GSM burst

4. Methodology, statistical features and machine learning

4.1 Overall workflow

The overall workflow for the extraction of the GSM burst, feature processing and extraction is provided in figure 3.

The workflow is composed by the following steps:

1. In the *Radio Frequency Signal Collection* phase, an Universal Software Radio Peripheral (USRP) Software Defined Radio platform is used to collect the signals in space. The test setup to capture the signals is described in section 4.2 and it is always the same for all the different mobile phones used in the test. The distance and antennas between the GSM mobile phone and the USRP at which the signals are collected is also described in section 4.2. The parameters and settings used for the collection of the signal in space are provided in table 1.

Sampling frequency	1 MSamples/sec full I/Q sampling
Front-end tuning frequency	890.2 MHz (GSM channel 1)
GSM Base station	Open BTS Version 3.1.3 on a USRP N200 platform

Table 1: Parameters for signal collection

Note that the base station and digitizer are fully GPS disciplined and synchronized using GPSDO. To support repeatability and stability, the same USRP digitizer as well as the same base station were used for all tests, all tests were performed after minimum half hour lock after the GPS receiver was properly synchronized on at least 4 GPS satellites. The use of different digitizers of different quality and the robustness of the identification algorithm is a topic for our future research.

2. In the *Baseband down conversion*, the signal is received and down-converted using a WBX, flexible frequency front-end compatible with the USRP with a passing bandwidth of 40 MHz and tuning capabilities from 20 MHz to 2 GHz. Then the signal is digitally down-converted by the USRP built-in DDC employing halfband and CIC (Cascaded integrator/comb) decimators from 100 MHz to 1MHz full IQ.
3. *Digital Filtering*. The base-band signal could have included interference from other sources in the environment and base band filtering is needed to ensure that only the uplink signal of the GSM phone under analysis is collected and processed as the test were all radiated and not conducted. We used a discrete-time, direct-form finite impulse response (FIR) filter using Parks-McClellan optimal FIR filter design. The Parks-McClellan algorithm uses the Remez exchange algorithm and Chebyshev approximation theory to design filters with an optimal fit between the desired and actual frequency responses. The filters are optimal in the sense that the maximum error between the desired frequency response and the actual frequency response is minimized. Filters designed this way exhibit an equiripple behavior in their frequency responses and are sometimes called equiripple filters. The parameters of the filter were following: a) Sampling Frequency = 1E6, b) Passband Frequency = 230000, c) Stopband Frequency = 320000, d) Passband Ripple = 0.17099735734, e) Stopband Attenuation = 0.001 and f) Density Factor = 720.
4. *Power normalization*. The power of the GSM uplink signal may not be the same for all the signal collections, given the variability of the propagation environment. As a consequence, in order to collect statistically significant samples, the RMS power of the signal was normalized.
5. *Extraction of GSM bursts*. This is one of the most critical part of the signal collection process because the bursts must be aligned and synchronized in a precise and consistent way. Only uplink traffic bursts from the DUT are kept. They are extracted from a single recording of 60 seconds and undergo an initial synchronization using thresholding at 0.05 of the normalized power. Empty bursts are discarded, traffic bursts are cut and inserted into a sliding window having the duration in samples of a GSM burst (655 symbols) + 10 in order to fully capture ramp-up and ramp-down. Bursts that are found outside the sliding window are re-synchronized, the non-compliant bursts are discarded and the sliding window is aligned again. This problem occurs due to the loss of samples in the digital section of the software radio.
6. *Burst Synchronization*. The collected bursts must be synchronized to ensure that sections of the GSM bursts used for fingerprinting are comparable, otherwise lack of synchronization could degrade the classification and validation procedures and reduce the overall accuracy of identification. This procedure discards the remaining bursts that are malformed to to sample loss in the digital section of the software radio. This

procedure utilizes the intrinsic correlation properties of the GSM mid-amble in order to achieve a synchronization error of one sample @ 1 MHz (10^{-6} seconds). This step is performed in order to compensate for *small out of synch bursts* which are caused by the mobile handset reacting to frequency and time alignment commands sent by the base station to the mobile phone. These corrections are performed by the base station and therefore should be considered as bias for signal classification purposes. The figure of the synchronized burst is shown in figure 4.

7. *Extraction of non-content burst segments.* Extraction of non-content burst segments. In this step, we extract the segments of the GSM burst which are not content related. This step is needed to avoid the fingerprint of the content (e.g., conversation on the GSM phone) rather than the GSM mobile phone. The extraction of the non-content sections requires the previous synchronization of the GSM burst to avoid the collection of content samples due to lack of synchronization among the bursts. The identification of the non-contents sections are based on the structure of the GSM burst as already described in section 3.
8. *Composition of mini-bursts.* This step has the objective to concatenate the non-content sections from the previous step in a new burst, which contains all the information needed for the classification and identification of the GSM mobile phone.
9. In the *Feature Extraction* step, various features are identified and applied to the mini-bursts (see section 4.3).
10. Then, application classification algorithms (e.g., SVM) are applied to the features extracted in the previous step. This is described in detail in section 4.4.

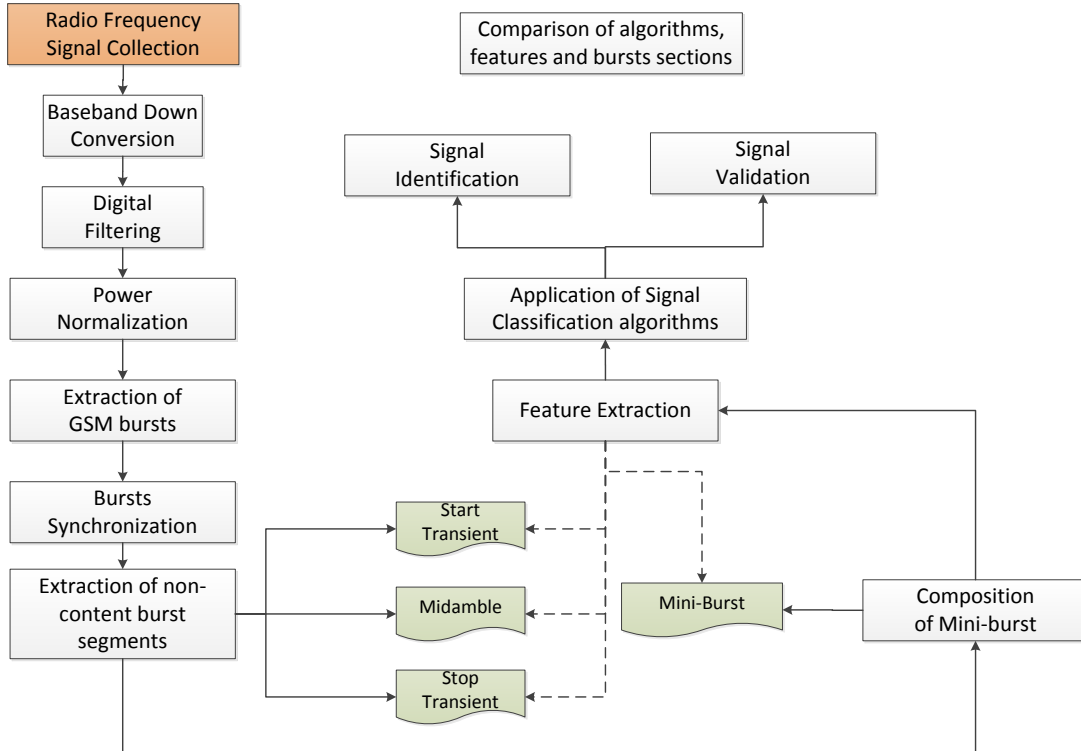


Figure 3: Workflow for the extraction of the fingerprints and classification

4.2 Experimental setup

The setup uses an Universal Software Defined Radio (USRP) system shown in figure 5. The measurement bench was set up in the laboratory room and consists of a table covered with

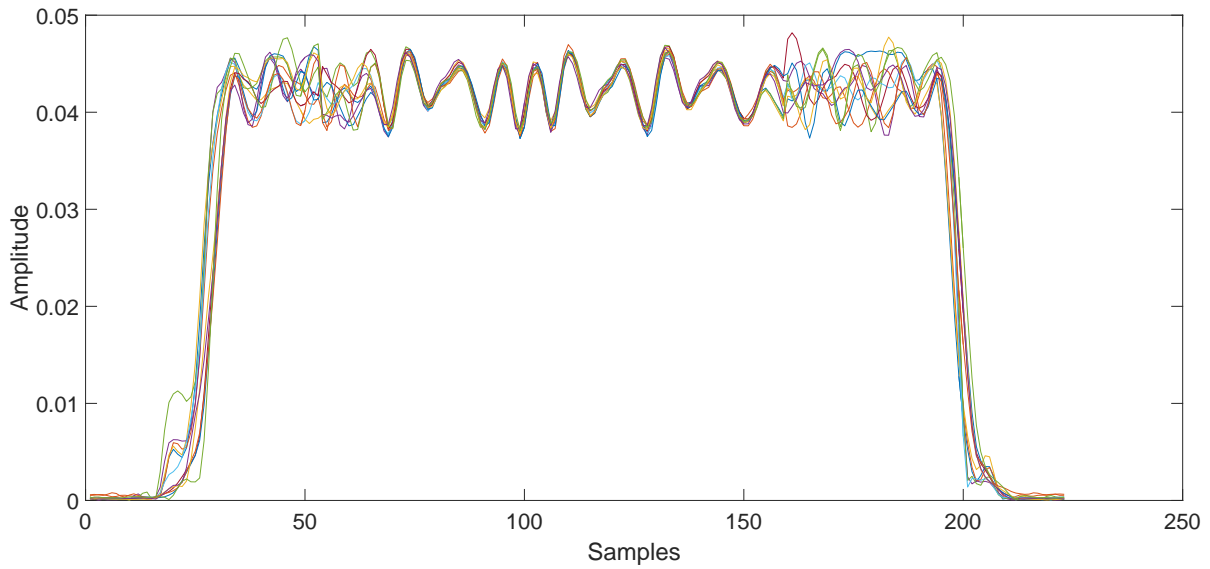


Figure 4: Synchronized GSM bursts

absorbers. The Device Under Test (DUT) and the measurement set are placed at the two ends the absorber layer at a distance of 84 cm. An image of the setup is presented in figure 5 and a schematic is presented in figure 6. The common parameters of the scenarios are presented in table 2 and the system timing parameters (specific to GSM) used for burst synchronization are presented in table 3.

A set of 12 phones were used in the experiment: 3 phones of HTC One model, 3 phones of iPhone model, 3 phones from Samsung and 3 phones from Sony Experia.

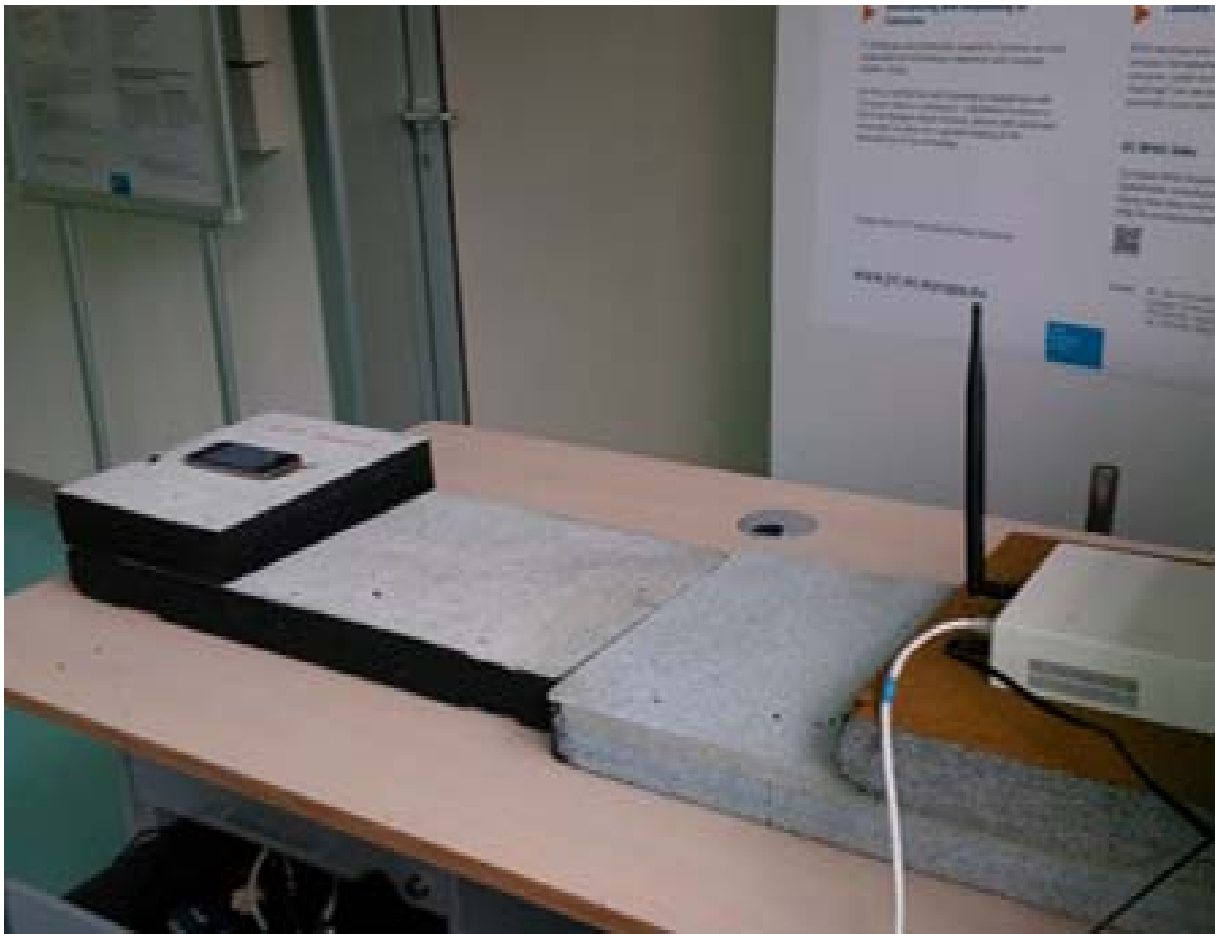


Figure 5: Image of the test bed for GSM mobile phone identification

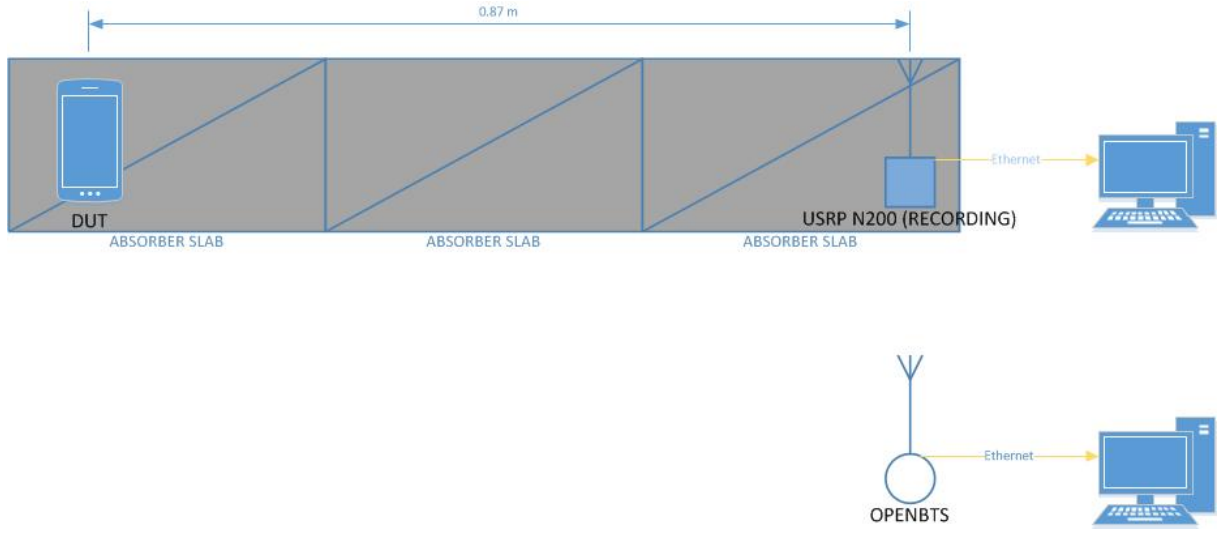


Figure 6: Test bed to collect the GSM signal

Sampling frequency	1 MS/sec IQ
Sample recording time	60 seconds
Downlink frequency	935.2 MHz
Uplink frequency	890.2 MHz
Synchronization	GPS using GPSDO (min 4 sat, min 30 min lock)
Distance between DUT and RX	0.84 m
USRP gain	5
GSM arfcn	1
OpenBTS version	3.1.3

Table 2: Experimental setup: test bed summary

Bit Rate	2.7086e+05 bps
Sample Frequency	1e+06 S/sec
bits per frame	1250 bps
frame duration	4.6150 ms
bits per burst	156.25 bits
samples per bit	3.6920 Samples
samples per frame	4615 Samples
samples per burst	576.875 Samples

Table 3: System Parameters used in the experimental setup

4.3 Statistical features

If we represent the complex sample signal of the GSM mini-burst in the time domain as:

$$S_{TD}(n) = I_{TD}(n) + jQ_{TD}(n) \quad (1)$$

we define the following features used for fingerprinting:

Mean

$$\mu = \frac{1}{N} \sum_{i=1}^N (S_{TD}) \quad (2)$$

Standard deviation

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (S_{TD} - \mu)^2} \quad (3)$$

Variance

$$\frac{1}{N-1} \sum_{i=1}^N (S_{TD} - \mu)^2 \quad (4)$$

Moment of Order 2, which is similar to the variance but with a divisor of N instead of N-1

$$\frac{1}{N} \sum_{i=1}^N (S_{TD} - \mu)^2 \quad (5)$$

Shannon Entropy, which is the expected value in average of the information contained in each message.

$$\text{ShannonEntropy} = - \sum_{i=1}^N (S_{TD}^2 * \text{Ln}(S_{TD}^2)) \quad (6)$$

Moment of Order 3

$$\sum_{i=1}^N (S_{TD}^3 - \mu^3) \quad (7)$$

Skewness, which is a measure of the asymmetry of the probability distribution of a random variable about its mean

$$\frac{1}{\sigma^3} \sum_{i=1}^N (S_{TD}^3 - \mu^3) \quad (8)$$

Kurtosis is a measure of how outlier-prone a distribution is.

$$\frac{1}{\sigma^4} \sum_{i=1}^N (S_{TD}^4 - \mu^4) \quad (9)$$

These features were applied to the entire mini-burst and to the specific section of the mini-bursts: the ramp up, the ramp down and the midamble. All these sections are not content related and they can be used for fingerprinting of the mobile device.

The application of statistical features to the GSM bursts make the fingerprinting process easier to process, because the algorithms only need to operate on the limited set of values of the statistical features rather than the bursts themselves (which are composed by a large number of samples). The challenge is to select the most appropriate set of statistical features for fingerprinting. This is achieved through a process called Sequential Feature Selection where the identification accuracy is calculated with a limited combination of features identified from set identified above. Then, a new feature is added to the combination of features to evaluate if the accuracy improves. The process continues until the accuracy reach an optimum value. At this stage the optimum combination of statistical features is identified. The identification accuracy can be based on different metrics defined in the next section where the machine learning algorithm is also defined.

4.4 Machine Learning algorithm

In the experimental study, we used the Support Vector Machine (SVM) with different Kernel function algorithms (see (12) for a description of SVM). The results shown in section 5 are based on the Gaussian Radial Basis Function kernel with different parameters but also other kernel functions have been used with SVM.

Support Vector Machine (SVM) is a supervised algorithm, which learns to classify the data points (e.g., originating from the observables), from the labeled training samples (e.g., the reference fingerprints). SVM separates the labeled set in two areas on a multi-dimensional surface by using a separating function, which can be of different type: linear, Radial basis function (RBF), polynomial, sigmoidal are the most common. Because the multi-dimensional surface is divided in two areas, SVM is a binary classifier and it can be directly used to distinguish between two mobile phones or for validation (to validate that the claimed identity of a mobile phone). See (13) for a detailed description of SVM. The extension of SVM to multi-classifier identification has been proposed by various authors (14)

and it is available in different libraries: the machine learning toolbox by MATLAB, LIBSVM and PRTools. The advantage of the SVM for fingerprint classification is that is well known for its high level of accuracy and robustness against outliers. SVM is less prone to overfitting than other methods (15). SVM is also quite efficient for binary classification, which is very important in the verification phase. The disadvantage is that SVM can be slow in the learning process and they can require a large amount of training time. Some algorithms used in SVM like Quadratic Programming (QP) methods can be computationally and memory intensive, so other Kernel methods should be preferred in fingerprint classification.

SVM is a binary classifier, and to perform classification of more than two systems as in our case (i.e., 12 mobile phones), we need to use a multi-classifier based on SVM. Two common approaches are the One Against One (OAO) and One Against All (OAA) techniques (12). OAA involves the division of an N (i.e., 12 in our case) class data-sets into N two-class cases, while OAO involves the creation of a classification machine composed by $N(N-1)/2$ machines for each pair of systems. While OAO is more computationally intensive than OAA ($N(N-1)/2$ against N), OAA has some disadvantages especially with unbalanced training data-sets. Then, we have decided to use OAO. The metric used to calculate the identification accuracy is based on the *confusion matrix*.

In the confusion matrix, each column of the matrix represents the instances of a predicted class while each row represents the instances of the actual class (and vice-versa). Because in our experiments we used 12 phones, the confusion matrix shown in the results section 5 has a dimension of 12×12 . In the confusion matrix, the correct guesses (i.e., true positive or negative) are located in the diagonal of the table, so it's easy to inspect the table for errors, as they will be represented by values outside the diagonal. The confusion matrix is also used in this report to define the accuracy. The metric is the sum of the diagonal values of the confusion matrix divided for all the values of the confusion matrix.

5. Experimental results

In this section, we provide the results of the application of the optimum set of statistical features and the machine learning algorithm. The confusion matrix of all the used phones is shown in figure 7. From the confusion matrix, we notice that it is very easy to distinguish mobile phones of different models, but it is possible to distinguish phones of the same model with less accuracy. This is understandable because mobile phones of the same model are built with the same materials and in the same manufacturing plant.

	HTC ONE	HTC TWO	HTC THREE	SAMSUNG ONE	SAMSUNG TWO	SAMSUNG THREE	SONY ONE	SONY TWO	SONY THREE	iPHONE ONE	iPHONE TWO	iPHONE THREE
HTC ONE	1285	136	578	0	1	0	0	0	0	0	0	0
HTC TWO	259	1560	180	1	0	0	0	0	0	0	0	0
HTC THREE	566	19	1410	4	1	0	0	0	0	0	0	0
SAMSUNG ONE	0	0	5	1568	285	142	0	0	0	0	0	0
SAMSUNG TWO	0	1	2	383	1049	565	0	0	0	0	0	0
SAMSUNG THREE	1	0	1	251	501	1246	0	0	0	0	0	0
SONY ONE	0	0	0	0	0	0	1250	328	177	70	73	102
SONY TWO	0	0	0	0	0	0	566	688	645	42	27	32
SONY THREE	0	0	0	0	0	0	331	556	1082	15	16	0
iPHONE ONE	0	0	0	0	0	0	80	36	19	1063	367	435
iPHONE TWO	0	0	0	0	0	0	68	13	10	567	954	388
iPHONE THREE	0	0	0	0	0	0	51	4	1	522	218	1204

Figure 7: Confusion matrix among all the phones used in the experiment

Then, we executed the main step of identifying a mobile phone against the reference library of already existing mobile phones.

The result is shown in figure 8. From the result one column vector, we can see that the system is able to recognize with very high accuracy the model of the phone, while it even recognizes the serial number with good accuracy.

In fact the inter-model accuracy is: $908/961 = 94.4\%$, while the average intra-model accuracy is 63.55% :

HTC ONE	0
HTC TWO	0
HTC THREE	0
SAMSUNG ONE	0
SAMSUNG TWO	0
SAMSUNG THREE	0
SONY ONE	33
SONY TWO	19
SONY THREE	1
iPHONE ONE	514
iPHONE TWO	908
iPHONE THREE	525

Figure 8: Selection vector for the recognition of a phone against the reference library

6. Conclusions

In this report, we have described the demonstration system used to implement the identification of a mobile phone on the basis of its RF wireless emissions. The demonstrator system can provide a very high accuracy for phones belonging to different models (inter-model identification) and limited accuracy for phones of the same model. Because counterfeit mobile phones are usually built with different electronic components than valid phones, the inter-model identification applies. We have shown that inter-model accuracy can reach a value of 94.4%, which is very high and it can be used for the detection of counterfeit mobile phones with great confidence. In addition, the identification of specific mobile phones can be used in forensics and criminal investigation.

References

- U. Guin, Ke Huang, D. DiMase, J.M. Carulli, M. Tehranipoor, and Y. Makris. Counterfeit integrated circuits: A rising threat in the global semiconductor supply chain. *Proceedings of the IEEE*, 102(8):1207–1228, Aug 2014.
- Ujjwal Guin, Daniel DiMase, and Mohammad Tehranipoor. A comprehensive framework for counterfeit defect coverage analysis and detection assessment. *Journal of Electronic Testing*, 30(1):25–40, 2014.
- Hui Zhou, Charles Nicholls, Thomas Kunz, and Howard Schwartz. Frequency accuracy & stability dependencies of crystal oscillators. *Carleton University, Systems and Computer Engineering, Technical Report SCE-08-12*, 2008.
- Smail Bachir, Nicusor E Calinoiu, and Claude Duvaud. New rf power amplifiers modeling and identification for wideband applications. *Analog Integrated Circuits and Signal Processing*, 83(2):161–172, 2015.
- H. J. Patel, M. A. Temple, and R. O. Baldwin. Improving zigbee device network authentication using ensemble decision tree classifiers with radio frequency distinct native attribute fingerprinting. *IEEE Transactions on Reliability*, 64(1):221–233, March 2015.
- Donald R Reising, Michael A Temple, and Michael J Mendenhall. Improved wireless security for gmsk-based devices using rf fingerprinting. *International Journal of Electronic Security and Digital Forensics*, 3(1):41–59, 2010.
- Christoph Böhm and Maximilian Hofer. *Physical unclonable functions in theory and practice*. Springer Science & Business Media, 2012.
- S. U. Rehman, K. W. Sowerby, S. Alam, and I. Ardekani. Portability of an rf fingerprint of a wireless transmitter. In *Communications and Network Security (CNS), 2014 IEEE Conference on*, pages 151–156, Oct 2014.
- Roger Clarke. Human identification in information systems: Management challenges and public policy issues. *Information Technology & People*, 7(4):6–37, 1994.
- Anil K Jain, Lin Hong, Sharath Pankanti, and Ruud Bolle. An identity-authentication system using fingerprints. *Proceedings of the IEEE*, 85(9):1365–1388, 1997.
- Gunnar Heine and Matt Horrer. *GSM networks: protocols, terminology, and implementation*. Artech House, Inc., 1999.
- Nello Cristianini and John Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.
- Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.
- Koby Crammer and Yoram Singer. On the algorithmic implementation of multiclass kernel-based vector machines. *Journal of machine learning research*, 2(Dec):265–292, 2001.
- Xindong Wu, Vipin Kumar, J Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J McLachlan, Angus Ng, Bing Liu, S Yu Philip, et al. Top 10 algorithms in data mining. *Knowledge and information systems*, 14(1):1–37, 2008.

List of abbreviations and definitions

BTS	Base Station
ETSI	European Telecommunications Standards Institute
FDMA	Frequency Division Multiple Access
GNSS	Global Navigation Satellite Systems
GSM	Global System for Mobile Communications
GPS	Global Positioning System
GPSDO	Global Positioning System Disciplined Oscillator
IC	Integrated Circuits
IPR	Intellectual Property Rights
LOS	Line of Sight
LTE	Long Term Evolution
MLE	Maximum Likelihood Estimator
OAA	One Against All
OAo	One Against One
PCM	Pulse Code Modulation
PUF	Physical Unclonable Functions
QP	Quadratic Programming
RF	Radio Frequency
RBF	Radial basis function
RSS	Frequency Difference of Arrival
RSSI	Received Signal Strength Indicator
SDR	Software Defined Radio
SNR	Signal to Noise Ratio
TDMA	Time Division Multiple Access
TDOA	Time Difference of Arrival
TOA	Time of Arrival
TS	Time slots
SVM	Support Vector Machine
UMTS	Universal Mobile Telecommunications System
USRp	Universal Software Radio Peripheral
VHF	Very High Frequency

List of figures

Figure 1. GSM burst definition from (11)	6
Figure 2. GSM burst	7
Figure 3. Workflow for the extraction of the fingerprints and classification	9
Figure 4. Synchronized GSM bursts	10
Figure 5. Image of the test bed for GSM mobile phone identification	10
Figure 6. Test bed to collect the GSM signal	11
Figure 7. Confusion matrix among all the phones used in the experiment	14
Figure 8. Selection vector for the recognition of a phone against the reference library	15

***Europe Direct is a service to help you find answers
to your questions about the European Union.***

Freephone number (*):

00 800 6 7 8 9 10 11

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

More information on the European Union is available on the internet (<http://europa.eu>).

HOW TO OBTAIN EU PUBLICATIONS

Free publications:

- one copy:
via EU Bookshop (<http://bookshop.europa.eu>);
- more than one copy or posters/maps:
from the European Union's representations (http://ec.europa.eu/represent_en.htm);
from the delegations in non-EU countries (http://eeas.europa.eu/delegations/index_en.htm);
by contacting the Europe Direct service (http://europa.eu/europedirect/index_en.htm) or
calling 00 800 6 7 8 9 10 11 (freephone number from anywhere in the EU) (*).

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

Priced publications:

- via EU Bookshop (<http://bookshop.europa.eu>).

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.



EU Science Hub

ec.europa.eu/jrc



@EU_ScienceHub



EU Science Hub - Joint Research Centre



Joint Research Centre



EU Science Hub

